

Visual and Thermal Data for Pedestrian and Cyclist Detection

Ahmed, Sarfraz, M. Nazmul Huda, Sujan Rajbhandari, Chitta Saha, Mark Elshaw, and Stratis Kanarachos

Author post-print (accepted) deposited by Coventry University's Repository

Original citation & hyperlink:

Ahmed, Sarfraz, M. Nazmul Huda, Sujan Rajbhandari, Chitta Saha, Mark Elshaw, and Stratis Kanarachos. "Visual and Thermal Data for Pedestrian and Cyclist Detection." In *Annual Conference Towards Autonomous Robotic Systems*, pp. 223-234.

https://dx.doi.org/10.1007/978-3-030-25332-5_20

Published as part of the Lecture Notes in Computer Science series, Vol 11650

ISSN 0302-9743

ESSN 1611-3349

Publisher: Springer

The final publication is available at Springer via http://dx.doi.org/10.1007/978-3-030-25332-5_20

Copyright © and Moral Rights are retained by the author(s) and/ or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This item cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder(s). The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holders.

This document is the author's post-print version, incorporating any revisions agreed during the peer-review process. Some differences between the published version and this version may remain and you are advised to consult the published version if you wish to cite from it.

Visual and Thermal Data for Pedestrian and Cyclist Detection

Sarfraz Ahmed¹, M. Nazmul Huda¹, Sujjan Rajbhandari¹, Chitta Saha¹, Mark Elshaw¹, and Stratis Kanarachos²

¹ School of Computing, Electronics and Mathematics, Coventry University, Coventry, CV1 5FB, U. K.

² School of Mechanical, Aerospace and Automotive Engineering, Coventry University, Coventry, CV1 5FB, U. K.

Abstract. With the continued advancement of autonomous vehicles and their implementation in public roads, accurate detection of vulnerable road users (VRUs) is vital for ensuring safety. To provide higher levels of safety for these VRUs, an effective detection system should be employed that can correctly identify VRUs in all types of environments (e.g. VRU appearance, crowded scenes) and conditions (e.g. fog, rain, night-time). This paper presents optimal methods of sensor fusion for pedestrian and cyclist detection using Deep Neural Networks (DNNs) for higher levels of feature abstraction. Typically, visible sensors have been utilised for this purpose. Recently, thermal sensors system or combination of visual and thermal sensors have been employed for pedestrian detection with advanced detection algorithm. DNNs have provided promising results for improving the accuracy of pedestrian and cyclist detection. This is because they are able to extract features at higher levels than typical hand-crafted detectors. Previous studies have shown that amongst the several sensor fusion techniques that exist, Halfway Fusion has provided the best results in terms of accuracy and robustness. Although sensor fusion and DNN implementation have been used for pedestrian detection, there is considerably less research undertaken for cyclist detection.

Keywords: Pedestrian Detection · Cyclist Detection · Sensor Fusion · Deep Neural Networks

1 Introduction

Recently, there has been a strong research focus on protecting vulnerable road users (VRUs), such as pedestrians, cyclists and motorcyclists [22]. This is because approximately half of the deaths on road accidents involve VRUs [34]. Of these VRUs, pedestrians and cyclists are at most risk as they do not have any form of special protection. And, although pedestrian detection techniques have been widely studied, cyclist detection has not received the same attention [10, 29]. This can be viewed as a concern as cyclists are also a part of the VRUs group, and can be just as susceptible to road traffic-related accidents. Also, it should be noted

that cyclists typically travel with higher speeds than pedestrians, which can lead to increased chances of road traffic-related incidents [29]. Therefore, when designing a detection system, both pedestrians and cyclists should be considered.

A number of approaches have been proposed for increasing the safety of pedestrians and cyclists. These approaches are becoming more relevant and necessary as autonomous vehicles are being used and tested on public roads. A major motivation for the development and advancement of autonomous vehicles has been to reduce the number of traffic accidents and thereby increase the safety of other road users [18]. To achieve this, various aspects of the design of the autonomous vehicle is being researched. One area that is being studied is sensor fusion. This is because the sensors are a vital part of the autonomous vehicle. They provide information, allowing the vehicle to sense its environment and detect objects within that environment. Using the information collected by the sensors, the vehicle is able to react and manoeuvre in an optimal and safe fashion [18]. By increasing the number of sensors, the perception of the vehicle can be increased in reliability and robustness.

Autonomous vehicles use various sensors, typically visual/thermal cameras, RADAR and LIDAR to sense its surrounding. For pedestrian and cyclist detection, visual sensors (i.e. cameras) are used as they produce high-resolution scene data [11]. Visual sensors have already been successfully applied for traffic sign detection and lane detection and are low cost to implement [21].

Pedestrian and cyclist detection have been a computer vision problem that has been widely explored due to its applications in autonomous vehicles, surveillance and tracking. Although there has been significant improvements made in the field of pedestrian/cyclist detection, it is yet to reach the levels to that of human perception [1, 24, 39]. The challenges that hinder the performance of these vision-based systems include occlusion, cluttered backgrounds and low-resolution images. Current detector designs use visual light spectrum cameras, which can provide high-resolution images in good lighting conditions. However, the quality of the images can be significantly reduced in poor lighting situations (i.e. night-time, evening) [24], decreasing the overall performance of the system. Visual (VS) and thermal-infrared (TIR) images have been fused for pedestrian detection to increase the robustness of the detection system [14, 24, 33].

Deep Neural Networks (DNNs) were implemented to further improve the effectiveness of general object detection, proving the effectiveness of DNNs for pedestrian detection [13, 24]. For example, a DNN-based detection technique yielded a mean average precision (mAP) of 53.7% when evaluated on PASCAL VOC 2010 in [13]. In a similar detection technique was proposed in [31], however, without the use of DNNs. The technique achieved an overall 35.1% mAP. However, it is still unclear which method(s) of fusion of sensor data that will provide the optimal results for pedestrian and cyclist detection using DNNs.

The primary purpose of this paper is determining the optimal method of sensor information fusion for pedestrian and cyclist detection based on DNNs. Although research in the field has been undertaken for pedestrian detection, there are relatively fewer studies for cyclist detection. Therefore, there is still a

need to study sensor fusion for cyclist detection as well as pedestrian detection. Based on previous works, DNNs should be able to provide more accurate results when compared to traditional methods of detection. The DNN is trained using multispectral data so that it can successfully extract features from both visible and infrared spectrums. This will aid in designing a detector that is more robust and accurate than the ones that currently exist for both pedestrian and cyclist detection. This will ultimately be a part of the prediction system for anticipating the behaviours of the pedestrian/cyclist for collision avoidance.

2 Pedestrian and Cyclist Detection

Although vision sensors have been widely researched for pedestrian detection in recent years, it remains a challenging problem that needs to be solved before human-like detection level is reached. The challenges are caused by the unpredictability of appearance of pedestrians due to occlusion, cluttered environment and pose. The cyclist detection problems are further complicated by multiple viewpoints of cyclist [22]. That is because cyclists, unlike pedestrians, can be perceived in a number of orientations. This causes difficulties as each orientation would have a different aspect ratio. The similarity between pedestrian and cyclists also poses another challenge. To overcome some of these challenges, a technique is proposed for detecting both pedestrians and cyclists [22]. The method uses the upper bodies (UB), which is similar for both pedestrians and cyclists.

3 Detection Systems for Autonomous Vehicles

It is vital that autonomous vehicles have a detection system that provides high levels of reliability while also being robust under various conditions (e.g. during both day and night, weather conditions). This is crucial to address safety concerns as more than 50% of accidents involving pedestrians occurs during night-time [2, 9]. To reduce this figure, thermal infrared (TIR) sensor-based cameras have been employed. Unlike visual (VS) light spectrum-based cameras, the TIR camera is not dependent on an external light source, rather it uses heat signatures of objects for detection and localisation. This makes TIR more effective at low light conditions, such as at night-time [2]. Both VS and TIR sensors will be discussed in the next sections. Whether using VS and TIR images together or independently, the process used for detection consists of at least feature extraction and classification. Some features that are collected during the daytime can be used for detection at night-time [2]. This is true when using the local binary pattern (LBP) technique, which was introduced for use with VS cameras. Features extracted using the LBP method only provide pixel intensity, which is the only channel of information for TIR cameras.

A common technique for feature extraction for pedestrian detection is the histogram of oriented gradient (HOG) [7]. The HOG technique extracts the gradient information for colour images and does not include information for the

intensity channel. HOG is widely used as it is robust. Considerable research was undertaken resulting in many variations of the approach. Thermal position intensity HOG (TPIHOG) works for the intensity information obtained via TIR sensors [2].

Once features are extracted, they are classified into classes. A commonly implemented classifier is the linear support vector machine (SVM) due to its decent performance and speed [37, 3, 28]. Several variations of SVM also exist. However, more recently, deep learning techniques have been used for general object detection as well as pedestrian detection. Classifiers based deep learning require large annotated datasets to be trained, which takes a longer time than other methods of classification [4]. However, deep learning improves accuracy and speed. The convolutional neural network (CNN) has also been widely adopted for pedestrian and cyclist detection.

4 Sensors

The detection of objects such as pedestrians, cyclists and vehicles are achieved through sensors. Sensors such as VS, IR, LIDAR and RADAR sensors, allow the autonomous vehicle to recognise objects in its environment. A single type of sensor approach cannot provide necessary levels of performance required of such a crucial and integral part of an autonomous vehicle [19] and therefore, sensor fusion can provide improved information. The focus of this study is fusion of VS and TIR sensors (i.e. cameras) used for pedestrians and cyclist detections.

4.1 Visible Light Sensors

Vision-based object detection has been widely studied, in particular pedestrian detection, but still remains a challenging task. VS cameras have been used with various methods of detection such as histograms of oriented gradient (HOG) with support vector machines (SVM) and adaptive boosting (AdaBoost) [32, 25, 7, 22, 30, 21]. The main concern using a VS-based pedestrian/cyclist detection is the performance susceptible to environmental conditions such as shadowing, uneven illumination, the decline in performance during low-light conditions (e.g. night-time, evening) and sensitivity to abrupt changes in illumination (e.g. going through a tunnel). They are also less effective due to occlusions caused by cluttered backgrounds or shape, size and pose of a pedestrian/cyclist. For this reason, multispectral (i.e. combined VS and TIR) data for sensor fusion have been introduced to overcome the shortcomings of the visible camera.

Even with these drawbacks of VS-based pedestrian detection, the majority of the research that has been undertaken has been based in the visible spectrum. This is due in part that the large datasets available are based on the visible spectrum [1]. These datasets are required for training DNNs to detecting desired objects.

4.2 Thermal Infrared Sensors

The spectral band of the TIR does not require an external light source, allowing it to overcome the shortcomings of the VS sensor. The far infrared (FIR) is the typical TIR sensor that is used for pedestrian detection [2]. The emitted radiation of human peaks in the FIR spectrum, making it an ideal TIR for pedestrian detection [40]. The images produced with FIR sensors are unaffected by variation in illumination. However, the performance of the thermal camera is affected by the high temperature [19]. For example, it is more difficult to detect a pedestrian or cyclist during hot days, as the background temperatures affect its resolution. Pedestrians and cyclists may also appear brighter or darker, depending on the outdoor temperatures [15]. It should be noted that FIR images include only one channel of information, which is intensity. This means that the images do not possess the same level of information provided though VS sensor cameras.

Thermal cameras have been successfully implemented for pedestrian detection [12, 20, 26, 27] with various detection methods such as HOG [20, 26, 23, 5]; shape and appearance-based detection [6] and contour saliency map (CSM) [8].

Visual and thermal data have been shown to be complementary [24], resulting in improved performance of the detection system in terms of accuracy and working conditions (i.e. different times of the day and under various weather conditions). To achieve an effective system, exploration of the ideal sensor fusion techniques is required.

4.3 Dual Sensor System

Combining two sensors could provide improved results whilst overcoming their individual drawbacks. With the effectiveness of TIR cameras for low-light conditions (i.e. night-time, evening) and the high level information provided by the visible cameras for well-illuminated environments.

A visible and thermal camera were combined for human tracking in an indoor environment [17]. The experiments, however, does not demonstrate the effects of outdoor influences, such as uneven lighting and shadowing. Night time pedestrian by combining visible and thermal cameras was demonstrated in [38]. This study, however, does not demonstrate the effects of high day temperatures on the thermal camera. A dual camera based on visible and thermal cameras for pedestrian detection that can function under various environmental conditions and during all times of the day was demonstrated in [19].

5 Sensor Fusion

Development of sensor fusion techniques for automotive purposes has been on the rise in recent years. This is because these techniques have been able to provide a higher level of accuracy in detection [35]. Multispectral pedestrian and cyclist detection can be one of three categories: pixel-level (Early Fusion) (see Fig. 1

(a)), feature-level (Halfway-Fusion) or decision-level (Late Fusion) (Fig. 1 (b)). Halfway Fusion is achieved by fusing the information after Early Fusion stage but before the Late Fusion stage. The categories are abstraction levels based on the layers of the DNN (typically a CNN-based approach) [33].

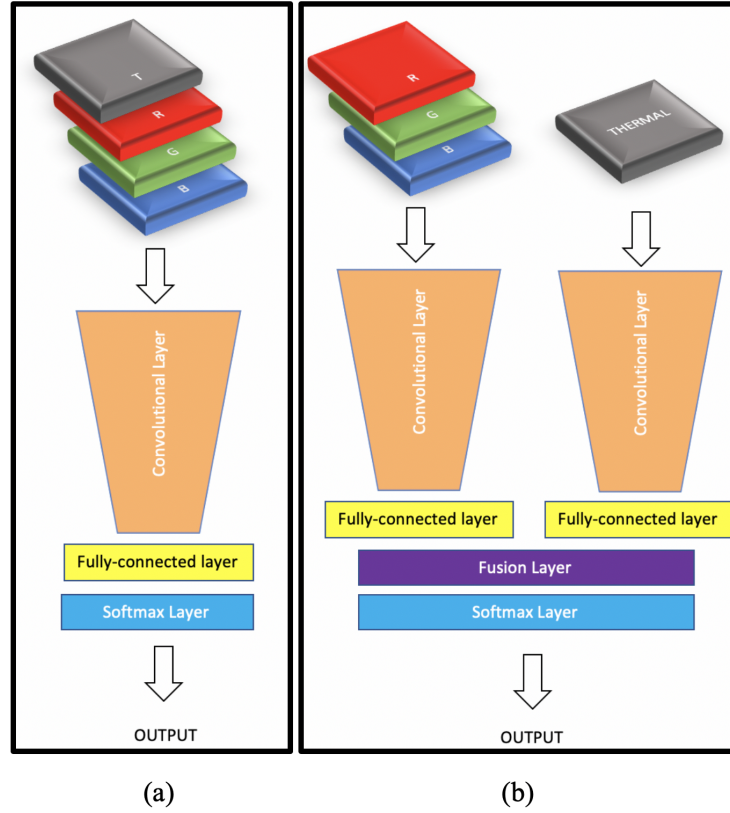
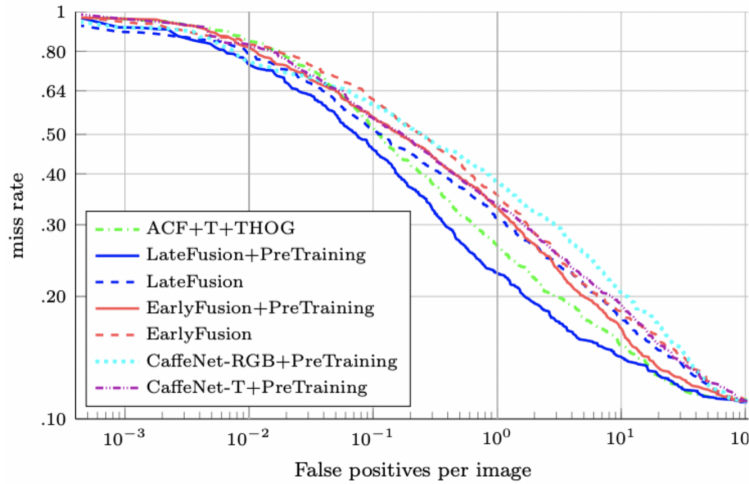


Fig. 1. Sensor Fusion using DNNs (a) Early Fusion (b) Late Fusion

Large amounts of labelled data is required for training of DNNs, which has been limited due to the costs associated with generating these large datasets. Therefore, pre-training using a large auxiliary dataset (such as ImageNet) has become rather popular [33]. This is because there is limited combined visible and thermal image datasets. Therefore, existing visible data was used for pre-training of the DNNs. In [2], the fusion techniques (with and without pre-training) is compared with ACF+T+THOG technique, a state-of-the-art solution for use of multispectral (visible and thermal) data [16] as a baseline. The LateFusion+PreTraining approach has shown to outperform the ACF+T+THOG baseline as well as the other fusion methods [33], see Fig.2. LateFusion+PreTraining

improved by approximately 6% and approximately 10% when compared to the baseline. Further, it illustrates that pre-training results in higher performance. It was also observed that the Early Fusion architecture was unable to improve on the baseline. It was speculated that it could be due to limited data in the KAIST dataset for the Early Fusion architecture without pre-training [33]. Overall, this means that Early Fusion approaches may not have learned meaningful features.



(a)

Detector	log-average miss rate		
	<i>all</i>	<i>day</i>	<i>night</i>
ACF+T+THOG	50.48 %	51.27 %	47.40 %
LateFusion+PreTraining	43.80 %	46.15 %	37.00 %
LateFusion	51.30 %	55.27 %	41.58 %
EarlyFusion+PreTraining	53.94 %	50.90 %	51.76 %
EarlyFusion	57.96 %	58.22 %	57.78 %
CaffeNet-RGB+PreTraining	56.52 %	53.51 %	63.40 %
CaffeNet-T+PreTraining	54.67 %	59.77 %	42.09 %

(b)

Fig. 2. Performance comparison of fusion techniques with ACF+T+THOG as a baseline and log-average miss rate during various times of the day (a) miss rate (b) log-average miss rate [33]

According to [24], the detector using fused data typically has enhanced performance when compared to a detectors based single sensor data. The single sensor approach, for example, vision sensors, may perform well during certain hours of the day, but that performance decreases as illumination decreases towards the evening and night-time. This is the case with some of the single sensor approaches in Fig. 3. (e.g. FasterRCNN-C and FasterRCNN-T).

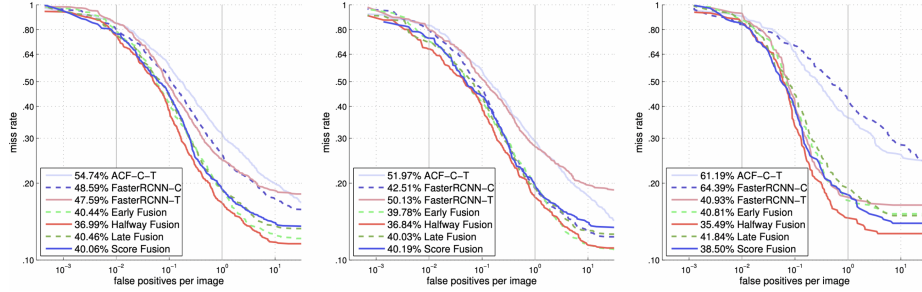


Fig. 3. Detection Results (MR/false positives per image) comparison with average of daytime and night-time (left), daytime (middle) and night-time (right) [24]

For the fusion models, a CNN-based was implemented for comparison of the performance of fused data and single data approaches (see Fig. 3). Four sensor fusion techniques were employed, Early Fusion, Late Fusion, Halfway Fusion and Score Fusion, in which Halfway fusion was the most effective technique. It had the lowest miss-rate (MR) (37%), 3.5% lower when compared with the other fusion methods. It is speculated that the Halfway Fusion technique was most effective as the various levels of the DNN is able to extract the most useful information from both VS and TIR data without reducing the high quality from the VS data [24]. This is not the case with Early Fusion (pixel-level), where low-level features are fused, features that may be irrelevant. Late Fusion and Score fusion meanwhile fuses only high-level features and confidence scores, respectively. This could make it difficult for the model to eliminate noise created during the fusion.

Overall the Halfway fusion outperformed the single sensor approaches, reducing the overall MR by approximately 11%. The approach is also more effective in low-light conditions and created fewer false alarms [24]. The experiment used the KAIST for evaluation, a dataset for pedestrian detection consisting of both visible and thermal spectra.

The study in [36], proposes the use of two DNNs for pedestrian detection based on visual and thermal data, the Cross-Modality Transfer CNN (CMT-CNN). The first DNN is used to learn the relationship between the RGB and thermal images. These learned relationships are then transferred to a second DNN for detection using RGB images only. This novel approach means that thermal data is not required for detection and is suitable for various illumination conditions. This method was compared to state-of-the-art pedestrian detection techniques on both the KAIST and the Caltech datasets.

As seen in Fig. 4, the CMT-CNN architectures performance is better than that of the best baseline (ACF-RGBT+HOG) for the KAIST dataset by approximately 5%. It should be noted that the CMT-CNN architecture only requires the RGB images, whereas the ACF-RGBT+HOG uses both RGB and thermal images. The approach also demonstrates that the approach would be very effective under low-light conditions.

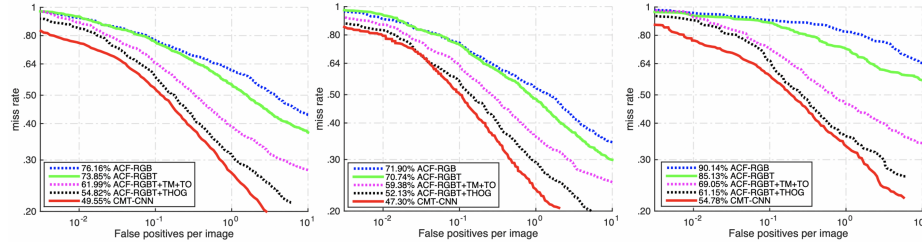


Fig. 4. Detection Results (MR/false positives per image) comparison with average of daytime and night-time (left), daytime (middle) and night-time (right) [36]

6 Conclusions

Although there has been a significant increase in research conducted for pedestrian detection using multispectral data, there is a considerable lack of work undertaken for cyclist detection using multispectral data. Sensor fusion is a challenge that has yet to be overcome for both pedestrian and cyclist detection. Therefore, future works should focus on sensor fusion for concurrent pedestrian and cyclist detection based on DNNs.

Majority of the research discussed in this paper found that the Halfway fusion technique provided the most effective empirical results when compared to a baseline detector. A state-of-the-art solution, ACF+T+THOG, has been used as a baseline to measure the performance of the sensor fusion techniques employed.

The Faster R-CNN seems to be used for pedestrian/cyclist detection due to its speed and reliability. Therefore, it would be sensible to use this architecture to evaluate sensor fusion techniques for pedestrian and cyclist detection. The KAIST dataset was utilised as it encompasses a large number of colour and thermal images in varying times throughout the day. It was also shown that an approach of pre-training a late fusion model can significantly outperform the baseline. However, early fusion fails in producing comparable results with the baseline. A reason for this failure could be that the approach could not learn significant abstract features from the multispectral data.

References

1. Angelova, A., Krizhevsky, A., Vanhoucke, V., Ogale, A., Ferguson, D.: Real-Time Pedestrian Detection With Deep Network Cascades. In: Proceedings of the British Machine Vision Conference. pp. 1–12 (2015). <https://doi.org/10.5244/C.29.32>
2. Baek, J., Hong, S., Kim, J., Kim, E.: Efficient pedestrian detection at nighttime using a thermal camera. *Sensors (Switzerland)* **17**(8) (2017). <https://doi.org/10.3390/s17081850>
3. Bertozzi, M., Broggi, A., Caraffi, C., Del Rose, M., Felisa, M., Vezioni, G.: Pedestrian detection by means of far-infrared stereo vision.

- Computer Vision and Image Understanding **106**(2-3), 194–204 (5 2007). <https://doi.org/10.1016/j.cviu.2006.07.016>
4. Biswas, S.K., Milanfar, P.: Linear Support Tensor Machine With LSK Channels: Pedestrian Detection in Thermal Infrared Images. *IEEE Transactions on Image Processing* **26**(9), 4229–4242 (9 2017). <https://doi.org/10.1109/TIP.2017.2705426>
 5. Chang, S.L., Yang, F.T., Wu, W.P., Cho, Y.A., Chen, S.W.: Nighttime pedestrian detection using thermal imaging based on HOG feature. In: *Proceedings 2011 International Conference on System Science and Engineering*. pp. 694–698. IEEE (6 2011). <https://doi.org/10.1109/ICSSE.2011.5961992>
 6. Congxia Dai, Yunfei Zheng, Xin Li: Layered Representation for Pedestrian Detection and Tracking in Infrared Imagery. In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops*. vol. 3, pp. 13–13. IEEE (2005). <https://doi.org/10.1109/CVPR.2005.483>
 7. Dalal, N., Triggs, B.: Histograms of Oriented Gradients for Human Detection. In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. vol. 1, pp. 886–893. IEEE (2005). <https://doi.org/10.1109/CVPR.2005.177>
 8. Davis, J., Sharma, V.: Robust detection of people in thermal imagery. In: *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004*. pp. 713–716. IEEE (2004). <https://doi.org/10.1109/ICPR.2004.1333872>
 9. European Road Safety Observatory: Traffic Safety Basic Facts 2012. Tech. rep., European Road Safety Observatory (2012)
 10. Gandhi, T., Trivedi, M.M.: Pedestrian protection systems: Issues, survey, and challenges. *IEEE Transactions on Intelligent Transportation Systems* **8**(3), 413–430 (2007). <https://doi.org/10.1109/TITS.2007.903444>
 11. Gerónimo, D., López, A.M., Sappa, A.D., Graf, T.: Survey of pedestrian detection for advanced driver assistance systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **32**(7), 1239–1258 (2010). <https://doi.org/10.1109/TPAMI.2009.122>
 12. Gilmore, E.T., D. Frazier, P., Chouikha, M.F.: Improved Human Detection Using Image Fusion. *Proceedings of the IEEE ICRA 2009* (2009)
 13. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. *EEE Conference on Computer Vision and Pattern Recognition (CVPR)* pp. 580–587 (2014)
 14. González, A., Fang, Z., Socarras, Y., Serrat, J., Vázquez, D., Xu, J., López, A.M.: Pedestrian detection at day/night time with visible and FIR cameras: A comparison. *Sensors (Switzerland)* **16**(6), 1–11 (2016). <https://doi.org/10.3390/s16060820>
 15. Hurney, P., Jones, E., Waldron, P., Glavin, M., Morgan, F.: Night-time pedestrian classification with histograms of oriented gradients-local binary patterns vectors. *IET Intelligent Transport Systems* **9**(1), 75–85 (2 2015). <https://doi.org/10.1049/iet-its.2013.0163>
 16. Hwang, S., Park, J., Kim, N., Choi, Y., So, I.: Multispectral Pedestrian Detection : Benchmark Dataset and Baseline. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* pp. 1037–1045 (2015)
 17. Jian Zhao, Cheung, S.c.S.: Human segmentation by fusing visible-light and thermal imaginary. In: *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops*. pp. 1185–1192. IEEE (9 2009). <https://doi.org/10.1109/ICCVW.2009.5457476>, <http://ieeexplore.ieee.org/document/5457476/>

18. Kocic, J., Jovicic, N., Drndarevic, V.: Sensors and Sensor Fusion in Autonomous Vehicles. In: 2018 26th Telecommunications Forum (TELFOR). pp. 420–425. IEEE (11 2018). <https://doi.org/10.1109/TELFOR.2018.8612054>, <https://ieeexplore.ieee.org/document/8612054/>
19. Lee, J.H., Choi, J.S., Jeon, E.S., Kim, Y.G., Le, T.T., Shin, K.Y., Lee, H.C., Park, K.R.: Robust pedestrian detection by combining visible and thermal infrared cameras. *Sensors (Switzerland)* **15**(5), 10580–10615 (2015). <https://doi.org/10.3390/s150510580>
20. Li, W., Zheng, D., Zhao, T., Yang, M.: An effective approach to pedestrian detection in thermal imagery. In: 2012 8th International Conference on Natural Computation. pp. 325–329. IEEE (5 2012). <https://doi.org/10.1109/ICNC.2012.6234621>
21. Li, X., Flohr, F., Yang, Y., Xiong, H., Braun, M., Pan, S., Li, K., Gavrilu, D.M., Flohr, F.: A new benchmark for vision-based cyclist detection. In: Proceedings of IEEE Intelligent Vehicles Symposium. pp. 1028–1033 (2016)
22. Li, X., Li, L., Flohr, F., Wang, J., Xiong, H., Bernhard, M., Pan, S., Gavrilu, D.M., Li, K.: A unified framework for concurrent pedestrian and cyclist detection. *IEEE Transactions on Intelligent Transportation Systems* **18**(2), 269–281 (2017). <https://doi.org/10.1109/TITS.2016.2567418>
23. Li, Z., Zhang, J., Wu, Q., Geers, G.: Feature Enhancement Using Gradient Saliency on Thermal Image. In: 2010 International Conference on Digital Image Computing: Techniques and Applications. pp. 556–562. IEEE (12 2010). <https://doi.org/10.1109/DICTA.2010.99>, <http://ieeexplore.ieee.org/document/5692620/>
24. Liu, J., Zhang, S., Wang, S., Metaxas, D.N.: Multispectral Deep Neural Networks for Pedestrian Detection. *British Machine Vision Conference* pp. 1–13 (2016)
25. Mikolajczyk, K., Schmid, C., Zisserman, A.: Human Detection Based on a Probabilistic Assembly of Robust Part Detectors. *Lect. Notes Comput. Sci* **3021**(3021), 6982 (2004)
26. Neagoe, V.e., Ieee, S.M., Ciotec, A.d., Bărar, A.p.: A Concurrent Neural Network Approach to Pedestrian Detection in Thermal Imagery. 9th International Conference on Communications (COMM), 2012 pp. 133–136 (2012). <https://doi.org/10.1109/ICComm.2012.6262539>
27. Olmeda, D., Armingol, J.M., de la Escalera, A.: Discrete features for rapid pedestrian detection in infrared images. In: 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems. pp. 3067–3072. IEEE (10 2012). <https://doi.org/10.1109/IROS.2012.6385928>, <http://ieeexplore.ieee.org/document/6385928/>
28. OMalley, R., Jones, E., Glavin, M.: Detection of pedestrians in far-infrared automotive night vision using region-growing and clothing distortion compensation. *Infrared Physics & Technology* **53**(6), 439–449 (11 2010). <https://doi.org/10.1016/J.INFRARED.2010.09.006>
29. Tian, W., Lauer, M.: Fast and Robust Cyclist Detection for Monocular Camera Systems. In: International joint Conference on Computer Vision Imaging and Computer Graphics Theory and Applications (VISIGRAPP) (2015)
30. Tian, W., Lauer, M.: Detection and Orientation Estimation for Cyclists by Max Pooled Features. In: Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications. pp. 17–26. SCITEPRESS - Science and Technology Publications (2017). <https://doi.org/10.5220/0006085500170026>

31. Uijlings, J.R.R., van de Sande, K.E.A., Gevers, T., Smeulders, A.W.M.: Selective Search for Object Recognition. *International Journal of Computer Vision* **104**(2), 154–171 (9 2013). <https://doi.org/10.1007/s11263-013-0620-5>
32. Viola, P., Jones, M.J., Snow, D.: Detecting Pedestrians Using Patterns of Motion and Appearance. *Proceedings of the 9th IEEE International Conference on Computer Vision* **1**(9), 734741 (2003). <https://doi.org/10.1109/ICCV.2003.1238422>
33. Wagner, J., Fischer, V., Herman, M.: Multispectral pedestrian detection using deep fusion convolutional neural networks. In: *European Symposium on Artificial Neural Networks* (2016)
34. World Health Organisation: *Global Status Report on Road Safety 2015 - Summary*. (2015)
35. Wu, T.E., Tsai, C.C., Guo, J.I.: LiDAR/camera sensor fusion technology for pedestrian detection. In: *2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*. pp. 1675–1678. IEEE (12 2017). <https://doi.org/10.1109/APSIPA.2017.8282301>
36. Xu, D., Ouyang, W., Ricci, E., Wang, X., Sebe, N.: Learning Cross-Modal Deep Representations for Robust Pedestrian Detection. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 4236–4244. IEEE (7 2017). <https://doi.org/10.1109/CVPR.2017.451>, <http://ieeexplore.ieee.org/document/8099934/>
37. Xu, F., Liu, X., Fujimura, K.: Pedestrian Detection and Tracking With Night Vision. *IEEE Transactions on Intelligent Transportation Systems* **6**(1), 63–71 (3 2005). <https://doi.org/10.1109/TITS.2004.838222>
38. Yuxi Chen, Chongzhao Han: Night-time pedestrian detection by visual-infrared video fusion. In: *2008 7th World Congress on Intelligent Control and Automation*. pp. 5079–5084. IEEE (2008). <https://doi.org/10.1109/WCICA.2008.4593753>
39. Zhang, S., Benenson, R., Omran, M., Hosang, J., Schiele, B.: How Far are We from Solving Pedestrian Detection? *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* pp. 1259–1267 (2016). <https://doi.org/10.1109/CVPR.2016.141>
40. Zhao, X., He, Z., Zhang, S., Liang, D.: Robust pedestrian detection in thermal infrared imagery using a shape distribution histogram feature and modified sparse representation classification. *Pattern Recognition* **48**(6), 1947–1960 (6 2015). <https://doi.org/10.1016/J.PATCOG.2014.12.013>